

Article

Historical collaborative geocoding

Rémi Cura^{1,†}, Bertrand Dumenieu^{2,3,†}, Julien Perret^{2,3}, Maurizio Gribaudo³

¹ IGN; remi.cura@gmail.com

² IGN; first.last@ign.fr

³ EHESS; last@ehess.fr

* Correspondence: julien.perret@ign.fr

† These authors contributed equally to this work.

Academic Editor: name

Version March 23, 2017 submitted to Humanities; Typeset by L^AT_EX using class file mdpi.cls

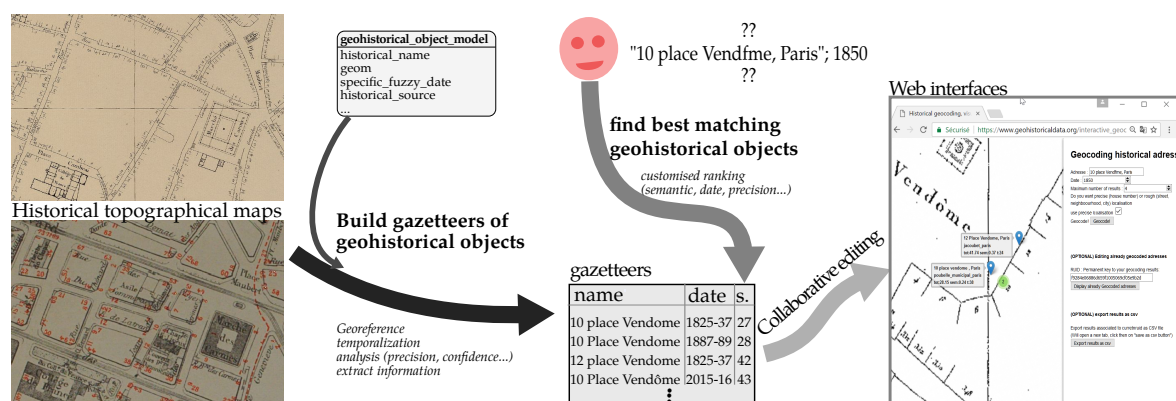


Figure 1. Graphical abstract: Gazetteers of geohistorical objects are created based on information extracted from georeferenced historical maps. Geocoding an historical address is then finding the best matching object based on customised function (semantic, temporal aspect, spatial precision, etc.).

Results can be displayed in a dedicated web interface for collaborative editing.

Abstract: The latest developments in digital have provided large data sets that can increasingly easily be accessed and used. These data sets often contain indirect localisation information, such as historical addresses. Historical geocoding is the process of transforming the indirect localisation information to direct localisation that can be placed on a map, which enables spatial analysis and cross-referencing. Many efficient geocoders exist for current addresses, but they do not deal with the temporal aspect and are based on a strict hierarchy (... , city, street, house number) that is hard or impossible to use with historical data. Indeed historical data are full of uncertainties (temporal aspect, semantic aspect, spatial precision, confidence in historical source, ...) that can not be resolved, as there is no way to go back in time to check. We propose an open source, open data, extensible solution for geocoding that is based on the building of gazetteers composed of geohistorical objects extracted from historical topographical maps. Once the gazetteers are available, geocoding an historical address is a matter of finding the geohistorical object in the gazetteers that is the best match to the historical address. The matching criteriae are customisable and include several dimensions (fuzzy semantic, fuzzy temporal, scale, spatial precision ...). AS the goal is to facilitate historical work, we also propose web-based user interfaces that help geocode (one address or batch mode) and display over current or historical topographical maps, so that they can be checked and collaboratively edited. The system is tested on Paris city for the 19-20th centuries, shows high returns rate and is fast enough to be used interactively.

Keywords: Historical dataset; geocoding; localisation; geo-historical objects; database; GIS; collaborative; citizen science; crowd-sourced; digital humanities

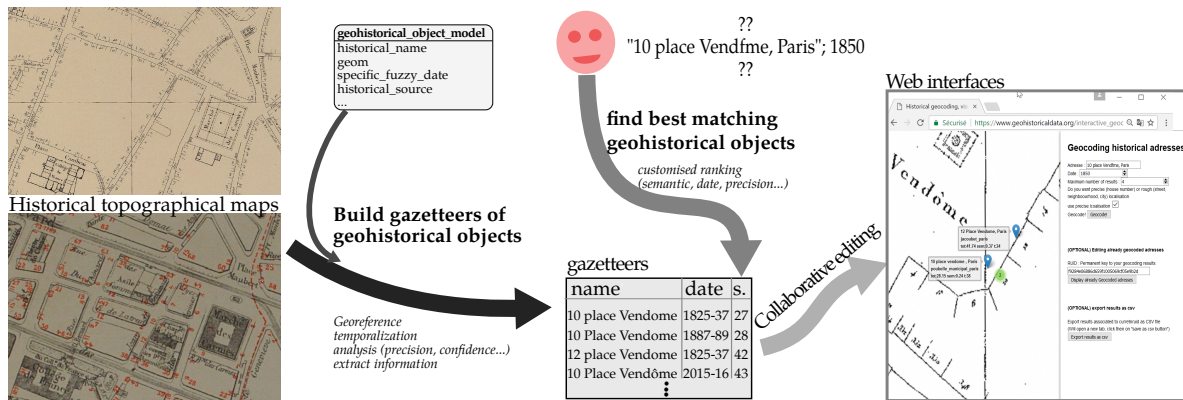


Figure 2. Graphical abstract: Gazetteers of geohistorical objects are created based on information extracted from georeferenced historical maps. Geocoding an historical address is then finding the best matching object based on customised function (semantic, temporal aspect, spatial precision, etc.). Results can be displayed in a dedicated web interface for collaborative editing.

1. Introduction

1.1. Context

The latest developments in *digital humanities* together with several important digitization efforts have provided humanists and social scientists with large data sets that can increasingly easily be accessed and used. In this article, we focus our attention on geohistorical data, which are historical sources conveying geographical information and thus allowing geohistorical studies. Such sources are diverse and may contain directly geographical information, such as maps, or indirectly geographical information such as addresses. The trouble is that historical data are incredibly diverse by their source, meaning and structure which implies that it has always been difficult to model them in a simple way. An address is here defined as "a description, including the names and any complementary pieces of information, which allows someone to uniquely identify a place" [1].

In order to use such addresses for spatial analysis, mapping or, more generally, for geohistorical studies, one needs to transform such descriptions to (geographical) coordinates. This process is called *geocoding*. Addresses found in historical sources might refer to addresses that have changed since (for many possible reasons) and might therefore not appear on recent (up-to-date) geographical databases. *Historical geocoding* thus requires a dedicated approach, dedicated tools and dedicated data and this article proposes several contributions in this direction.

1.2. Related work

The *geocoding* process answers a recurring need in research and in several application domains such as navigation. As such, it is quite well documented in the literature [2,3] and is commonly provided as a service (with *Nominatim*¹ or *Gisgraphy*² for instance). As previously shown [2], *geocoding*

¹ nominatim.openstreetmap.org

² www.gisgraphy.com

tools can be characterized in terms of their main components: input and output data, reference dataset and processing algorithm. The input is the locational description the user wants transformed into geographical coordinates. For instance, such an input could be "13 rue du temple, Paris, France". The reference dataset contains geospatial objects associated to address components. The processing algorithm consists in finding the best match between the latter components and the input description. Finally, the output usually contains a geographical entity and a qualification of the match (perfect match or approximate for instance).

These tools answer millions of geocoding queries each day with great success. Indeed, geocoding services quality can be estimated via two very important criteria [4]. The first is the database quality: how complete and up to date is the database? The second is result characterization: how spatially precise is the result and what is the associated confidence?

Despite their quality, such geocoding approaches can not be used for (geo)historical data for three main reasons. The first is that existing geocoding services do not take into account the temporal aspect of the query or the datasets they rely on. Indeed, they usually rely on current data, such as *OpenStreetMap*³ data, continuously updated. As such, they implicitly work on a valid time that is the present (or possibly the interval between the start of the database construction and now). The second reason is that they rely on a complete, strongly hierarchical database which is verifiable (i.e. there is always a way to check the real localization of an address). On the opposite, historical data are not directly verifiable: one has to check (possibly incomplete and conflicting) available (geo)historical sources and, very often, to make assumptions or hypotheses. Such hypotheses are continuously challenged and updated by new discoveries, and there is no way to give a truly definitive answer. Indeed, primary sources can also be wrong or misleading. The third reason is that historical sources available to construct a geocoding database are sparse (both spatially and temporally), heterogeneous, and complex. We believe that all these specificities call for a dedicated approach.

Similar observations have already been made in the context of archival data by the UK National Archives [5]. Large historical event gazeteers already exist [6] and provide an important basis for the construction of the reference dataset. We found few related articles ([7],[8]), and for all of them geocoding was not a focus.

We could not find an historical geocoding method that consider the geocoding result characterization. Yet this aspect is essential for historical geocoding because of the very unprecise and sparse nature of geohistorical data. Indeed, geocoding results have to be validated and/or edited manually. Considering the large amount of addresses (> 100000 *for Paris*) and the potential complexity of the task, this is clearly a lot of work. Fortunately, several projects such as *OpenStreetMap* have lead the way for what is usually called *Volunteered Geographical Information* (VGI) [9] of *crowdsourcing geospatial data* [10]. This approach consists in using a collaborative approach to solve the problem collectively, usually by implying citizens in the process. As suggested in a recent typology of participation in citizen science and VGI [11], different levels of participation can be defined. These levels go from "crowdsourcing", where the cognitive demand is minimal, to "extreme citizen science" or "collaborative science", where citizens are involved in all stages of the research (problem definition, data collection and analysis). In the rest of this article, we propose a collaborative historical geocoding approach that opens a way for a simpler participation of citizens in geohistorical research (thanks to dedicated interactive tools), but also for a more collaborative geohistorical science (thanks to a reproducible research approach [12–15], open source tools and open data).

1.3. Approach and contributions

In this article, we focus on the historical geocoding problem. Following the classical approach, we will present in particular the construction of a geohistorical database and the development of

³ openstreetmap.org

matching (data linkage) methods that fully use the temporal aspects of the geohistorical data and the input query.

Our main contributions are:

- a formalisation of the historical geocoding problem,
- a minimal model of historical and geo-historical objects that can be easily re-used and extended,
- an open source geocoding tool that is powerful, easy to use and can be extended with any geohistorical data,
- a graphical tool to control and edit the geocoding results, which can then optionally be used to enrich the geohistorical database,
- qualifications of geocoding results in term of semantic, spatial, temporal aspects.

2. Methods

Based on historical sources and historical topographical maps, we extract geohistorical objects that are collected into gazetteers. These (geo) historical object are modeled in a generic way into a Relational Database Management System (RDBMS). Geocoding an historical address is then finding the geohistorical object in the gazetteers that best match this historical address, which is done via various distances that can be customised by the user. Last the results can be displayed via a web interface over current or historical topographical maps, and further checked and edited collaboratively. These edits are then integrated into the geocoder database.

2.1. Building gazetteers: Extracting geohistorical objects from historical topographical maps

The starting point to build gazetteers is information we extract from historical topographical maps. The first part of extraction is to digitize the maps and georeferencing the map in a pre-defined geographic coordinate system. This maps are historical sources, and as such an historical analysis is performed to estimated the probable valid time (temporalization), spatial accuracy, completeness, confidence, relation to other historical maps, etc. The whole process is carefully designed and explained in detail in [16]. Then geo-historical objects are extracted from the referenced historical maps, manually (in a collaborative way), or with the help of computer vision.

2.1.1. General consideration about building a spatio-temporal database

Extracting information from topographical historical maps amounts to building a spatio-temporal database. There are several approaches to do so, and we stress that we do not attempt to create a continuous spatio temporal database. Instead, we store representations of the same space at multiple moments in history, the well-known snapshot model [17]. The main advantages is that for a given moment in time we can have several conflicting snapshots coexisting. This is essential, as solving the conflict may not be possible (no way to check by going back in time), and reporting this several conflicting geocoding result to historian may help appreciate these results. The drawbacks of this model, *i.e.* information redundancy and its inability to store the changes themselves, can be overcome during the geocoding process.

2.1.2. Historical topographic maps as geohistorical sources

Usually, these snapshots are built on archival - texts and maps - or archaeological sources which provide information about the spatial organization of an area. In our approach, we focus on historical topographic maps as the main sources for two main reasons:

- the way they portray spatial information is close to modern day topographic mapping, making the integration of the information they convey in a GIS easier;
- the main goal of topographic maps is to provide a reliable depiction of geographical objects and their arrangement.

Although this choice seriously reduces the number of possible sources and therefore lessens the quantity of accessible spatial information, it aims at efficiency. Indeed, topographic maps are a good compromise between their reliability, the quantity of spatial information they contain and the complexity of extracting these informations.

In our model, a snapshot is constructed for each historical topographic map by relying on three processes:

1. georeferencing the map in a pre-defined geographic coordinate system,
2. assigning the map a *valid time*,
3. extracting geographical objects from the map.

2.1.3. Georeferencing topographic historical maps

We have to establish a correspondence between each pixels of the historical maps and modern geographical coordinates. To do so, we first choose a common spatial reference system (SRS). Then we identify common geographic features between historical maps and current maps (matching points). Last, we compute a warping transform that will respect at best the matching points.

One issue can be that finding matching points between current maps and historical maps can be increasingly difficult as we go back in time, because there are less and less un-ambiguous matching points. Consider for instance the city of Paris, where the French revolution and its consequences combined with the 19th century transformations by Haussmann resulted in massive changes in the shape of the city. To this end, we can start by georeferencing e.g. 20th maps to current maps, then georeference e.g. 19th maps to 20th maps, and so one for even older maps.

Common spatial reference system (SRS)

The choice of a SRS is not easy, as each SRS induces projections errors that depends on the covered area. Whatever the choice of SRS, it is essential that the implied accuracy is well-known and documented in order to qualify the absolute accuracy of each geo-referenced map. We restrain ourselves to SRS using meters as base units (opposed to SRS using degrees), as they are much closer in nature to those used in historical topographical maps.

Feature pairs selection

Feature pairs identification is a critical step because the number, distribution and quality (i.e. positional accuracy, reliability, confidence) of the features strongly influence the quality of the georeferencing. While the quality of selected features depends on each map, a simple rule of thumbs is to select as much as possible homogeneously distributed feature pairs [18].

To achieve a satisfactory feature pairs selection, three parameters have to be considered: the geometric type of the features, their nature and the method used to identify them.

The most classical geometry of feature pairs is 2D points; lines or surfaces may also be used, and possibly even curves, in a similar spirit to [19].

On historical maps, mapping themes can accuracy can vary greatly (e.g. building vs forest) due to purpose of the map, or making process. Then the features pairs should be selected in the same type. Optionally, we can rely on geodetic features drawn on the map such as meridian or parallels provided we can fully characterised the geodetic characteristics of this lines.

The actual identification of feature pairs can be achieved by automatic or manual processes. Automatic approaches are notably used for historical aerial photographs, where feature detection and matching algorithms are well fitted [20]. Common GIS tools offer georeferencing tools allowing to manually select pairs of *ground control points* that corresponds to features identified in both the input and the reference maps. Such tools are often the basis of historical maps georeferencing because: (1) they are easy to utilize and (2) they allow historians to control the quality and reliability of the identified points using co-visualization between both maps.

Choosing a geometric transformation model

Once an acceptable set of paired features has been identified, the last step is to compute the transformation from the input map to the reference. Several transformation models have been proposed through years: global transforms (affine, projective), global with local adaptations (polynomial-based) and local transforms (rubbersheeting, kernel-based approaches). Studies have been conducted to assess the relevance of these transformation for historical maps [18,21,22]. They show that choosing a model is mostly a matter of compromise between the final spatial matching between the feature pairs (i.e. the expected residual error) and the tolerable distortion of the map regarding its legibility. Exact or near-perfect matching between feature can be achieved with local transforms and high order polynomials, whereas the internal structure of the map is most preserved by global transformations. Low order polynomials offer a compromise between both constraints.

2.1.4. Temporalization: locating geohistorical sources in time

Georeferencing is a way of locating multiple maps in the same reference space. Similarly, *temporalization* is the process of locating each geohistorical source in time. When building spatio-temporal snapshots from historical maps, the key problem is to determine the moment where the map is representative of the actual state of the area it portrays, i.e. the *valid time* of the map. We considered the valid time of each map to be the period starting with the beginning of the topographic survey and ending with the publication of the map, which are often uncertain. Representing uncertain or imprecise periods of time is a common issue when dealing with historical information and many authors relied on the fuzzy set theory to represent and reason on imperfect temporal knowledge [23,24].

We model imprecise valid times as trapezoidal fuzzy sets, that is a trapezoidal function of time between 0 (the source provides no information at this time) and 1 (geographical entities portrayed in the map are regarded as existing and tangible at this time). We rely on the pgSFTI⁴ postgres extension.

For instance, figure 3 illustrates the *valid time* of a map whose topographic survey started in year 1775, ended between 1779 and 1780 and which was engraved late 1780.

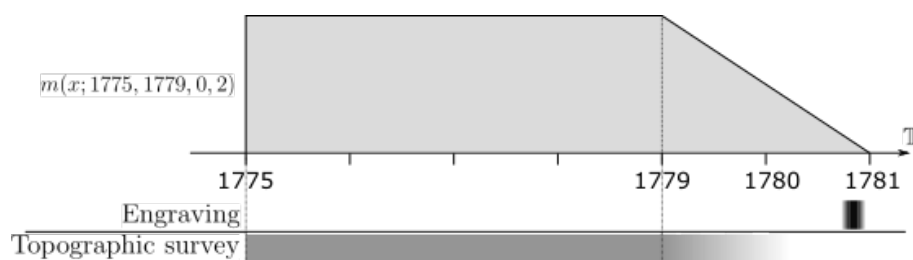


Figure 3. An uncertain valid time modelled as a trapezoidal fuzzy set function

2.1.5. Extracting information from maps

When historical maps have been georeferenced and temporalized, extraction can be extracted to create the basis for geohistorical objects.

The most classical way to extract information from maps is by human action with a classical GIS software (e.g. QGIS). However one historical map of Paris contains a large amount of information to be extracted (e.g. tens of thousands of street names, hundreds of thousands of building number, ...)

⁴ <https://github.com/OnroerendErfgoed/pgSFTI>

A first solution is then to use computer vision and machine learning methods to create automatic extraction tools. These tools can process the whole map in few hours. Regrettably such tools are difficult to design, are very specific to one historical map, and may produce low quality results.

Recently collaborative approaches have proved to be very efficient for building big geographical databases in a relatively short period (OSM⁵, NYPL⁶).



Figure 4. In this example, hand written text is automatically detected and extracted (red) from an historical map.

2.2. Modelling geohistorical objects

Information extracted from historical maps is used to create gazetteers. Those are made of geohistorical objects.

To this end, we design a geohistorical objects model with all necessary attributes and also flexibility to adapt to the great variety of geohistorical object types and sources.

Our goal is to provide an universal minimal (geo)historical object model that can be used by other and easily extended when necessary.

2.2.1. modelling choices

Geohistorical data are extremely diverse, both in term of historical sources and in term on how the sources were dealt with by historian. // As such, historian use complex tailored models.

modelling approach

We do not aim at modelling all the geohistorical data in all their complexity. Instead, we propose to model the bare minimal common properties of all geohistorical objects, and offer mechanisms so this model can be easily extended and tailored to the specificities of the data.

To define the bare minimal model, we start from the very nature of a geohistorical object, that is both an historical object and a geospatial object. The extension mechanism is provided via a database-object oriented design using table inheritance, and is packaged into a PostgreSQL extension⁷.

⁵ <http://www.openstreetmap.org>

⁶ <http://buildinginspector.nypl.org/>

⁷ https://github.com/GeoHistoricalData/geohistorical_objects

2.2.2. geo-historical objects model

Geo-historical objects have both an historical and a geospatial part. We stress that modelling historical source and numerical origin process of a geohistorical object is an essential part. The detail of the model are illustrated in figure The figure 5.

Historical aspect

In its essence and historical object is defined by names, sources, and temporalization.

- **Names.** By names, we mean the historical name that was used to identify the object in the object historical context, and the current name that is used by historian to identify the object in the current context.
For instance, the Eiffel tower in Paris historical name may be "tour de 300 mètres", but today it is referenced as "tour Eiffel".
- **Sources.** A historical object is defined by an historical source (document), where the object is referenced. Beside the historical source, the way the object was digitized in this source is also essential.
For instance, a street name may have the Jacoubet topological map as historical source, and would have been digitized via collaborative editing on the georeferenced map.
- **Temporalization.** Any historical source is associated with temporal information (fuzzy dates), which is the period during which the source is probably relevant. Beside the hisorical source temporal informatioin, a geohistorical object can also have its own temporal information.
For instance, a street may have been extracted from a historical map having been drawn between 1820 and 1842. Besides this information, using other historical documents allow to narrow the probable existence of this streets to 1824-1836.

Geospatial aspect

A geo-historical object is also defined by geospatial information: the object geometry, and the object precision model.

- **Geometry.** An object has a geometry which follows the OGC standard⁸. It may be a point, polyline, polygon, or a composition of any number of those, in a specified SRS. Geometry can then be transformed into a common SRS if need be and be used jointly. The geometry is extracted from the historical source (in a manual or automatic way)
- **Spatial precisions.** The geospatial object sources have spatial precision information. This precision express the spatial uncertainty of the historical source (the person drawing the map may have made mistakes) and the spatial uncertainty of the digitizing process (the person editing the digitised map may have made a mistake). One historical source may contain several precision, one for each geohistorical object type.
For instance, an historical map may contain building and roads. Buildings may have a different spatial precision (5 metres) than road axis (20 metres). Besides, the digitising process precision may have been of 5 metres.

Formal model

2.2.3. A database of geohistorical objects

We defined the model for a geo-historical object, which is based on two names, two sources, fuzzy dates and a geometry. This defines the core of a generic geohistorical object.

⁸ <http://www.opengeospatial.org/standards>

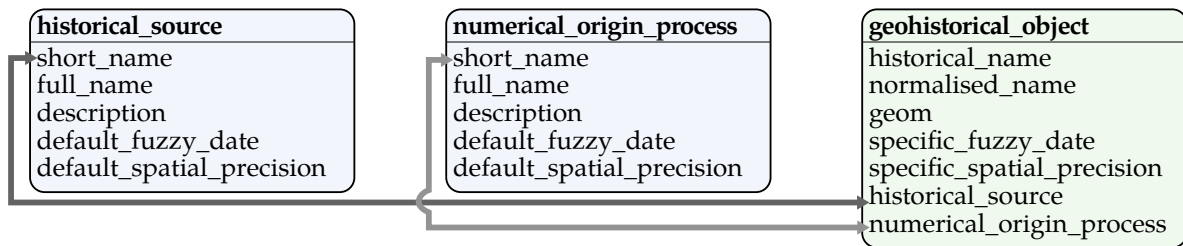


Figure 5. Geohistorical object model

Yet this geohistorical object model is easily extendible using the table inheritance mechanism, an object-oriented design mechanism that is available in PostgreSQL (see Figure 6).

Table inheritance

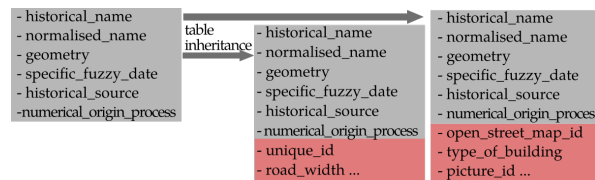


Figure 6. Example of table inheritance: columns are herited, and others can be added

The concept of table inheritance is simple. When a table *child* is created as inheriting from a table *parent*, *child* will have at least the columns of *parent*, but can also have other columns (provided there is no name/type collision).

This means in our case that a table of geohistorical objects will inherit from the main geohistorical object table, i.e. will have all the core columns of geohistorical objects (names, sources, temporal aspect, spatial aspect), but can also have its own tailored column, providing the necessary flexibility.

Another key aspect of table inheritance is that the *parent* table is queried, the query will be executed on not only the rows of *parent* table, but also on the rows of all *child* table. This means that all tables using the geohistorical object model will be virtually grouped and accessible from one table.

Simulated inheritance of index and constraints

The PostgreSQL table inheritance mechanism is however limited in some aspects, because constraints and index can not be inherited.

Constraints are essential, because they are used to guarantee that any geohistorical object will correctly use existing sources from the sources tables ("historical_source" and "numerical_origin_process").

Indexes are also essential, because when using hundred of thousand of geohistorical objects, they are needed to help speed the queries.

We index not only names, but all geohistorical object core columns (names, sources, temporal aspect, spatial aspect). We propose a registering function that the user can execute only once when creating a new geohistorical object table.

Modelling a geohistorical object from the user perspective

The practical steps to create geohistorical objects are simple.

1. Add the historical source and numerical origin process in the source and process tables.

2. Create a new table inheriting geohistorical objects and containing your additional custom columns
3. Use the registering function with this table name
4. Insert your data in the table.

2.3. Geocoding historical addresses with geohistorical object gazetteers

In the previous section we explained how we create gazetteers of geohistorical objects from maps.

1. Historical map is scanned.
2. Scan are georeferenced using hand picked control points.
3. Historical work allow to estimate temporal information and spatial precision of the map.
4. Roads name and axis geometry is extracted from the scan (manually or automatically).
5. Building number is extracted from the scan (manually or automatically).
6. In some cases, building number can be generated from the available data.
7. normalised names are created from historical names.
8. Geohistorical objects are created.

The next step is to use these gazetteers to geocode historical addresses.

2.3.1. Historical geocoding concept

In our method, geocoding something is finding the most similar geohistorical objects within the available gazetteers, which then provides the geospatial information.

This approach relies on two key components: gazetteers of geohistorical objects, and a metric to find the best matches.

This approach allows to perform geocoding in a broad sense, as it does not rely on a structured address (number, street, city ...), but rather on a non constrained name.

2.3.2. Creating geohistorical object gazetteers for geocoding

Geohistorical object gazetteers are key for the geocoding. These objects are extracted from topographical historical maps and inserted into geohistorical objects tables. Each table form a gazetteer.

Database architecture for geocoding

We again use the PostgreSQL table inheritance mechanism. To this end, we create two tables dedicated to geocoding. Now gazetteers tables that will be used in geocoding must inherits from these two tables. "precise_localisation" table is for building number geohistorical objects, e.g. "12 rue du temple, Paris". "rough_localisation" table is for road axis, neighbourhood, cities geohistorical objects.

We chose to have two separate tables for ease of use and performance.

Geocoding queries are then performed on the two parents tables, but thanks to inheritance, these parents tables virtually contains all the gazetteers table containing the actual geohistorical objects, as illustrated in figure 7.

2.3.3. Finding the best matches

Once geohistorical objects gazetteers describing precise and rough localisation are available, geocoding is finding the best match between the input query and the objects.

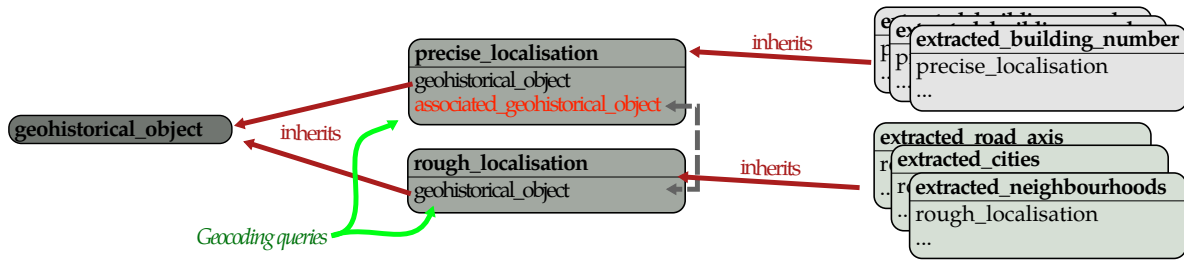


Figure 7. Geocoding architecture.

Concept

We call the potential matches "candidates", and the problem is then to rank the candidates from best to worst. The user can choose how much candidates he wants, depending on the application. For an automated batch geocoding, the best match (top candidate) is optimal. For a human analysis of data, several matches may be more interesting (top 10 candidates for instance).

What can be qualified as "best" depends on the user expectations. We provide a number of metrics that can be combined by a user into a tailored ranking function. The function is expressed in SQL, with access to all postgres math functions. We describe the available metrics and give example of such function.

Example

For instance when a user geocodes the address "12 rue de la vannerie, paris" in 1854, user may be more interested into geohistorical objects that are semantically close (e.g. a geohistorical object "12 r. de la vannerie Paris", 1810), or maybe geohistorical objects that are more temporally close (e.g. "12 r. de la Tannerie Paris", 1860).

Metric: semantic distance w_d

We use the semantic distance provided by PostgreSQL Trigram extension (pg_trgm⁹), which compare two strings of characters by comparing how much successive set of 3 characters are shared. For instance "12 rue du temple" will be farther away from "12 rue de la paix" than from "10 r. du temple".

Metric: temporal distance t_d

Both the address query and the geohistorical object contains fuzzy dates. We design a simple fuzzy date distance that behaves relatively familiarly. We cast fuzzy dates as geometry (polygon), where the x axis is the time, and the y axis is the probability of existence of the object. Then the distance from date A to date B is $\text{shortest_line_length}(A,B) + \text{Area}(A) - \text{Area}(A \cap B)$. Note that this distance is asymmetric.

Metric: building number distance b_d

To get building number distance, a function tries to extract the building number both from the input address query (b_i) and from the geohistorical object (b_d). If b_i and b_d have same parity, the distance is $|b_d - b_i|$. If parity is different, the distance is $||b_d - b_i| + 10|$. In France, building numbers have in general the same parity on each side of the street (e.g. Left: 1,3,5,..; Right: 2,4,6..).

⁹ <https://www.postgresql.org/docs/current/static/pgtrgm.html>

We analysed current building number in Paris and determined that on average, given a building number b_i , the closest building number with a different parity has a 10 number difference.

Metric: spatial precision s_p

Another way to rank the geohistorical objects candidates is to use their spatial precision. The spatial precision of a geohistorical object is either the object specific spatial precision when it exists, or the default spatial precision of this object source.

Metric: scale distance s_d

The geocoding architecture can provide localisation at different scale, depending on the user requirement. For instance if the user scale of study is the city, there is no need to perform a more precise geocoding. Therefore the user can specify a target scale range (S_l, S_h). Then given a geohistorical object whose geometry is buffered ($geom_b$) with its spatial precision, the scale distance is defined by $least(|\sqrt{area(geom_b)} - S_l|, |\sqrt{area(geom_b)} - S_h|)$. The formula $\sqrt{area(geom_b)}$ gives an idea of the spatial scale of the geohistorical object.

Metric: geospatial distance g_d

The user may provide an approximate position for the area he is interested in. For instance in France both city "Vitry-le-François" (East) and "Vitry-sur-Seine" (near Paris) exist, but are very spatially far away. A user expecting results in the Paris area may provide a geometry (a point for instance) near Paris. Then the classical geodesic distance is computed between the provided geometry and the candidates geohistorical object.

Example of matching function

The different metrics can be weighted and combined depending on the user needs. The equation 1 gives an example that favour good semantic similarity, but not at the price of big temporal distance.

$$100 * w_d + 0.1 * t_d + 10 * n_d + 0.1 * s_p + 0.01 * s_d + 0.001 * g_d \quad (1)$$

2.4. Collaborative editing of geohistorical objects

The geocoding approach we have presented in the previous section works inside a PostgreSQL database. Given an input address and fuzzy date, plus a set of parameters, it returns the geohistorical objects that matches the most the input.

Yet the geocoding results are only as good as the gazetteers are. The geohistorical objects within the gazetteers may be spatially un-precise, mistakenly named or simply missing. Given that the volume of geohistorical objects is large (for Paris, approximately 50 k building number per historical map), we create a collaborative platform to facilitate geocoding, visualising the results and editing the geospatial objects when necessary.

To this end, we create a dedicated web application so collaborative editing is possible without having to install specific tools.

2.4.1. About collaborative editing

Given that the geocoding method we use is open source and the data produced is open data, the collaborative approach makes sense. [25]

2.4.2. Collaborative editing architecture

The figure 8 outlines the architecture used for collaborative editing.

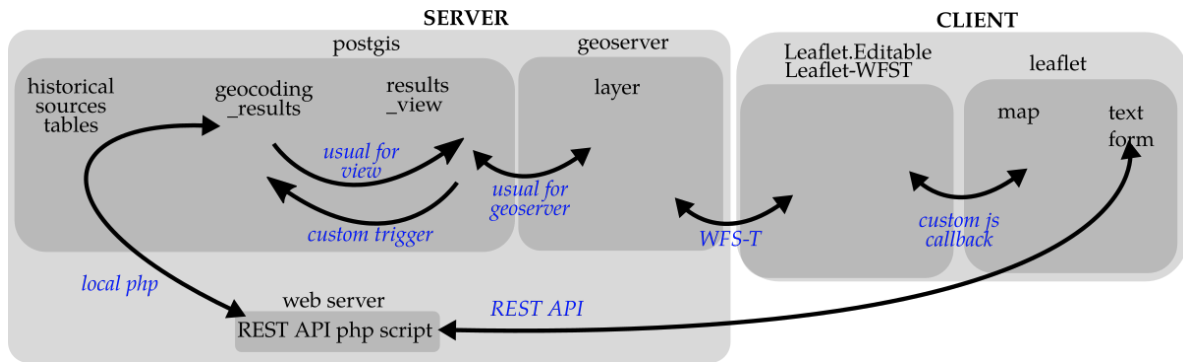


Figure 8. Conceptual architecture for interactive display and edit of geocoding results.

Architecture

The heart of the architecture is the PostgreSQL database server, which contains the geohistorical objects gazetteers that will be used for geocoding as well as geocoding function.

A webserver can geocode address and return results via a REST API.

However, the webserver has another option where the results are not returned, but instead written in a result table along with a random unique identifier (RUID). The RUID is then the key that permit to display and edit the results. To this end, a geoserver can access (read and edit) the result table via the WFS-T protocol.

A web application based on Leaflet acts then as a UI to display and edit the results via the geoserver.

Persistence of geocoding results and edits

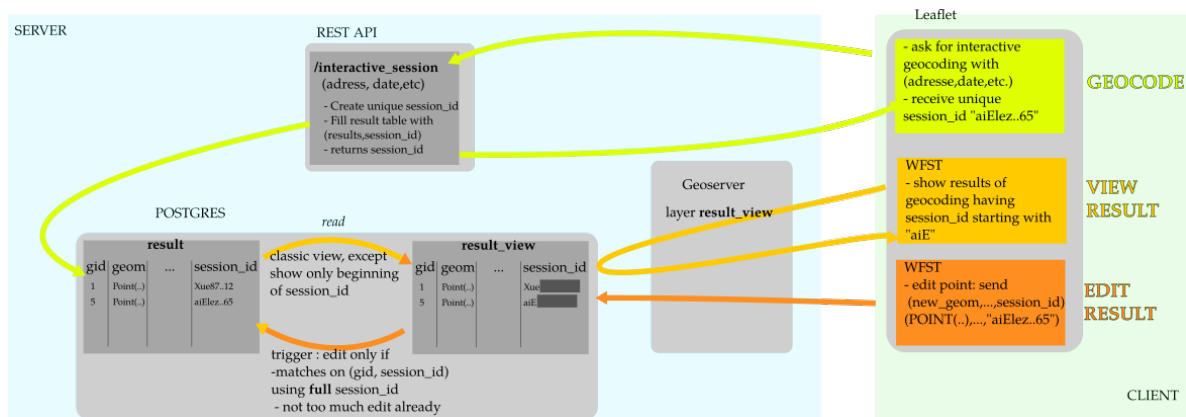


Figure 9. Using a secret key to allow collaborative display and edit

The architecture that allows persistence of results is illustrated in figure 9. When using the RUID mechanism, each geocoding result (that is the found geohistorical object from the gazetteers) is associated to this RUID. That way the user can always access its results, regardless of the computer session or browser cache issues.

To edit, a specific mechanism is used. The user does not directly edit the result table, as he could potentially edit other people results. Instead, the user edit a dedicated result_view that acts like a bouncer. It allow edit only if the edit is occurring on a row that has the user RUID.

User edit of the geospatial objects do of course not affect the source data, for a tracking purpose.

Instead an user edit automatically create an edited copy of the geohistorical object in a dedicated table "user_edit_added_to_geocoding" that is a gazetteers and is used by the geocoding process. In this table are inserted the edited geohistorical object. The objects retain their "historical_source", but their "numerical_origin_process" is changed to properly document the fact that they are the result of a collaborative editing.

2.4.3. Collaborative editing user interface

We consider that building efficient user interface is very important for historical geocoding. In particular, many end users are specialised on history rather than on computer science, and thus an easy access to geocoding is essential.

All our interfaces are web-based for a maximum of compativity. We propose three interfaces whose results are shared.

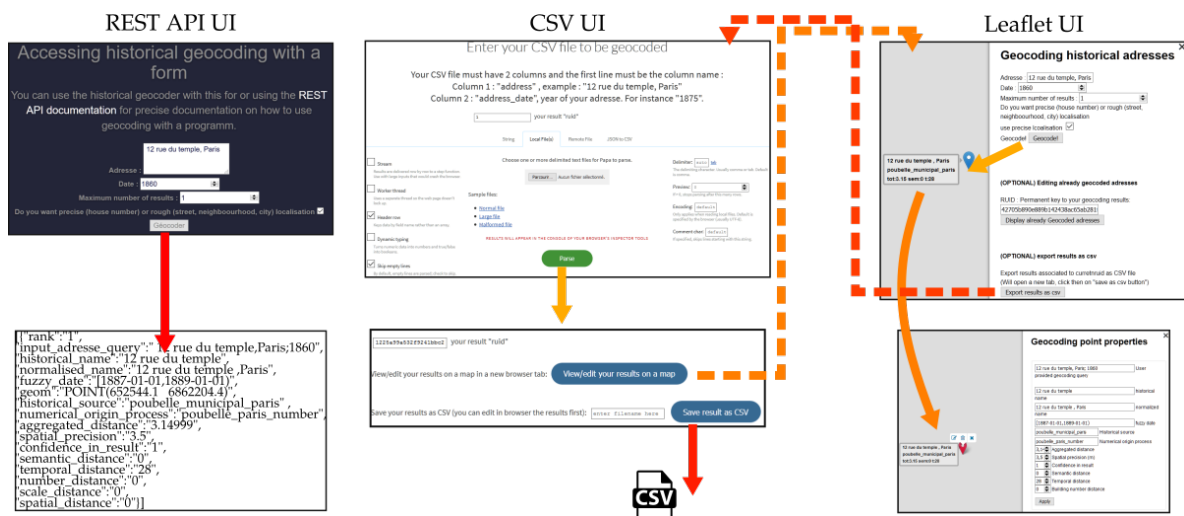


Figure 10. Various Web User Interface for use of historical geocoding.

Interface for REST API.

The simplest interface we propose is a form that helps build the necessary REST API parameters. Indeed, REST API works via URL containing precise parameters, and it can be tedious to manipulate. For instance: "https://www.geohistoricaldata.org/geocoding/geocoding.php?adresse=2012 rue du temple, Paris&date=1860&number_of_results=1&use_precise_localisation=1"

This interface is designed to be used in an automated way, for batch geocoding.

Interface for batch geocoding via CSV files.

In our experience historian often work with spreadsheet files, where each line will be a potential historical object, along with an address and a date. To facilitate the geocoding of these addresses, we propose an User Interface that can read Coma Separated Value (CSV) files (which is a standard spreadsheet format), and geocode the address and date within. This interface is build around PapaParse¹⁰ Javascript framework.

Then the results of geocoding can be either downloaded as a CSV file, or displayed and edited in a web application.

¹⁰ <http://papaparse.com>

Interface for display and edit of results.

The most complex interface we propose is based on Leaflet¹¹ Javascript framework. There, the user can geocode an address, or use already geocoded address via the RUID mechanism (see Section 2.4.2), be it from previous sessions or from geocoded CSV files. The geocoding results are displayed on top of a relevant historical map, and can be edited. User can edit results geometry as well as results names (historical and normalised). We stress that although such edit are stored in the database, and used by further geocoding queries, they do not affect source data, by design.

3. Results

We perform several experiments to validate our approach. First we use the geohistorical model to integrate objects extracted from historical topographical maps from the 19th century for the city of Paris, and the current OpenStreetMap road axis and building numbers for Paris city surroundings. We successfully integrate the road axis, building numbers, and neighbourhoods to the geocoder sources.

We then perform multiscale geocoding of dozens of thousand of historical addresses extracted manually by historian and extracted automatically by automatic process.

Last we test the collaborative editing of geohistorical object in two scenarios: analysis (several results for one address), and edit (efficiency of check/edit top results for several addresses).

3.1. Geohistorical objects sources

We mainly use three historical sources of geohistorical objects to perform geocoding. The first two are Historical topographic maps of Paris from the 19th century. These maps are georeferenced then street axis (and possibly building numbers) are manually extracted. The third historical sources are road axis and building number for Paris surrounding extracted from current Open Street Map data.

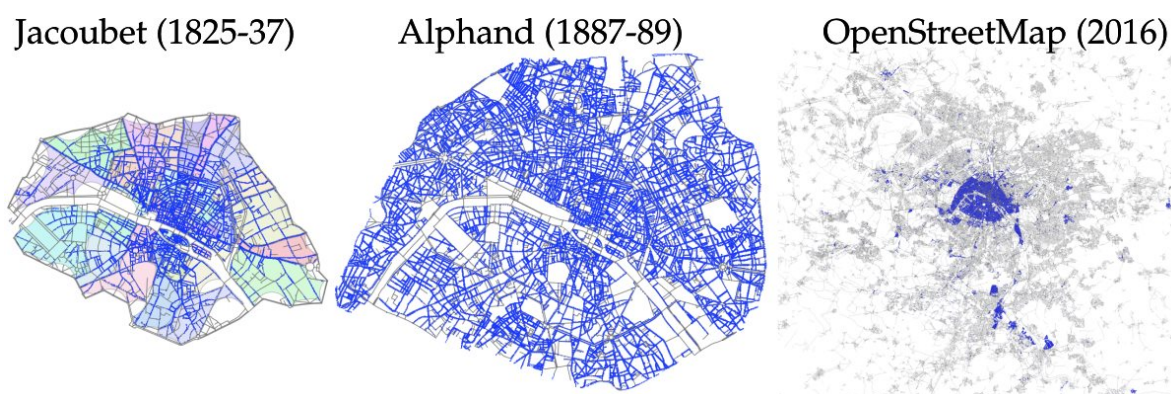


Figure 11. geohistorical objects used from geocoding extracted from the source maps.

3.1.1. Historical topographic maps used

We integrated two major French atlases of Paris from the 19th century as geohistorical sources. The first one is the "Atlas municipal de la Ville, des faubourgs et des monuments de Paris"¹² created at the scale of 1 : 2000 between 1827 and 1836 by Theodore Simon Jacoubet, an architect who was working for the municipal administration of Paris. The second atlas is the 1888 edition of the

¹¹ <http://leafletjs.com>

¹² Municipal atlas of the city, suburbs and monuments of Paris.

"Atlas municipal des vingt arrondissements de la ville de Paris"¹³. For readability reasons, we refer to the first atlas as the "Jacoubet atlas" and the second as the "Alphand atlas"¹⁴. The Jacoubet atlas depicts a city standing between the housing development following the sale of the properties confiscated during the French Revolution and the majors changes in the urban structure arising from the emergence of the first train stations in 1837-1840 and the Haussmannian transformations.

The Alphand atlas is a portrayal of Paris at the scale of 1 : 5000, after most of the Haussmannian transformations (major rework of Paris urbanism in the 19th century) and after the city was merged with 11 of its neighboring municipalities in 1860. Both atlases contain large scale topographic views of Paris, separated in several sheets (54 and 16 respectively) and portray the urban street network with each street named, building of public purposes and religion buildings (see figure 12). In addition, the house numbers are specified for most of the streets in the city, although the Alphand atlas pictures only the numbers at the start and end of each street section. Both atlases are also built upon triangulation canvas covering the entire city, allowing us to expect a high positional accuracy of the geographical objects they contain.

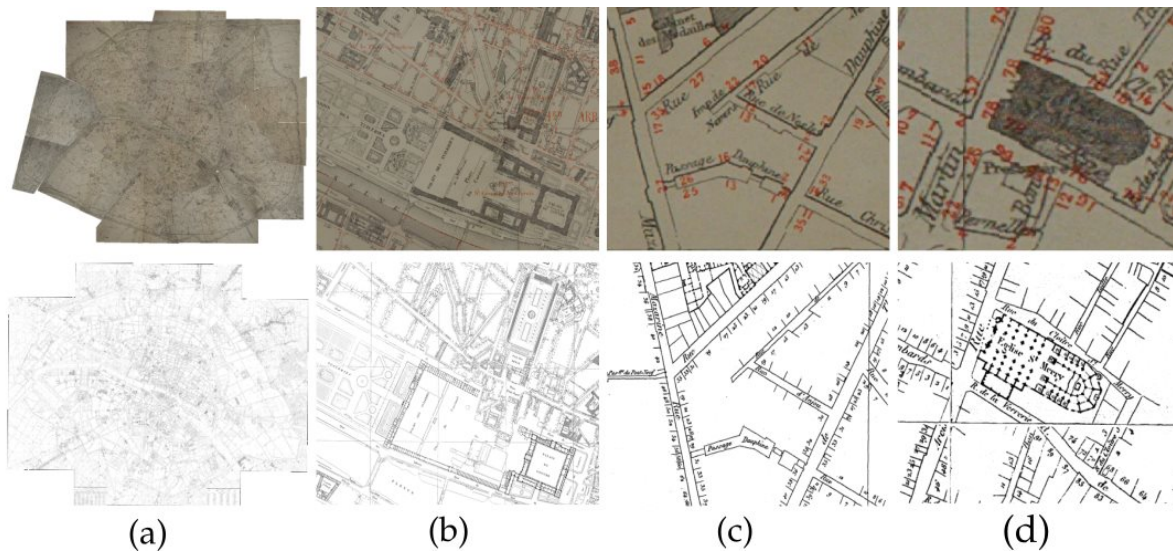


Figure 12. Samples of the georeferenced Alphand Atlas (2nd row) and Jacoubet Atlas (1st row) at different scales: whole city (a), district (b) and urban islet (c). Column (d) shows how buildings are portrayed in the maps.

We georeferenced the two atlases using the grids drawn on the maps, which are aligned on the Paris meridian, as a pseudo-geodetic objects to identify feature pairs. The dimensions of the grid cells also appear on the maps, allowing us to reconstruct the grids in a geographic reference system. We have chosen to georeference the maps in the Lambert I conformal conic projection, which uses the Paris meridian as prime meridian and rely on the NTF (Nouvelle Triangulation Française) geodetic datum. The main advantage of this projection is that it is locally close to the planar triangulation of Paris used in the atlases. Thus, the projection of the maps can be reasonably approximated by the Lambert I projection, making the reconstruction of the grids in the target coordinate reference system straightforward. In addition, since both maps are at high scale and are reliable because they are official maps with high positional accuracy, we used rubbersheeting as the geometric transform model. The georeferencing process applied for each atlas was the following:

¹³ Municipal atlas of the 20 districts of Paris

¹⁴ From the name of Jean-Charles Alphand who was at the time the director of the department of public works of Paris.

- reconstruct the meridian-aligned grid with Lambert I coordinates;
- in each sheet, mask the non-cartographic parts out (cartouche, borders, etc.);
- for each sheet, set pairs of ground control points at each intersection between the vertical and horizontal lines of the grids in the map and in the reconstructed grid;
- transform each sheet with a rubbersheeting transform based on the ground controls points previously identified on the grids.

geohistorical objects extraction

Based on these atlases, vectorial road axis are manually drawn and the road name inputted. For Alphand map, the building number at the beginning and end of each street segment is also inputted. For Jacoubet, the building numbers from a previous map (Project Alpage, Vasserot map, [26]) are adapted to fit the Alphand map. Multiple series of successive checking and editing are performed using ad hoc visualisations and tools.

For Alphand, building numbers are then generated based on the available information (for each street segment, for each side, beginning and ending number) by linear interpolation, and an offset. The size of the offset is estimated by using current Paris road width when the road has not changed too much.

3.1.2. Other geohistorical sources

We also use current data from OpenStreetMap. We use the version of the data that has been transformed to be used by the Nominatim geocoder. Custom scripts extract road axis and building numbers. The dataset covers Paris city and its surroundings, and is dated to 2016.

3.2. Geocoding of Historical datasets

One of the end goal of our geocoding tool is to be useful for historians. Therefore, we contacted several historians working on Paris (19th century). They had been collecting historical addresses, which we geocoded by importing their data into the geocoding server.

The following figure shows an extract of the thousands of geocoded addresses, while the table gives an overview of the number of success and timing.

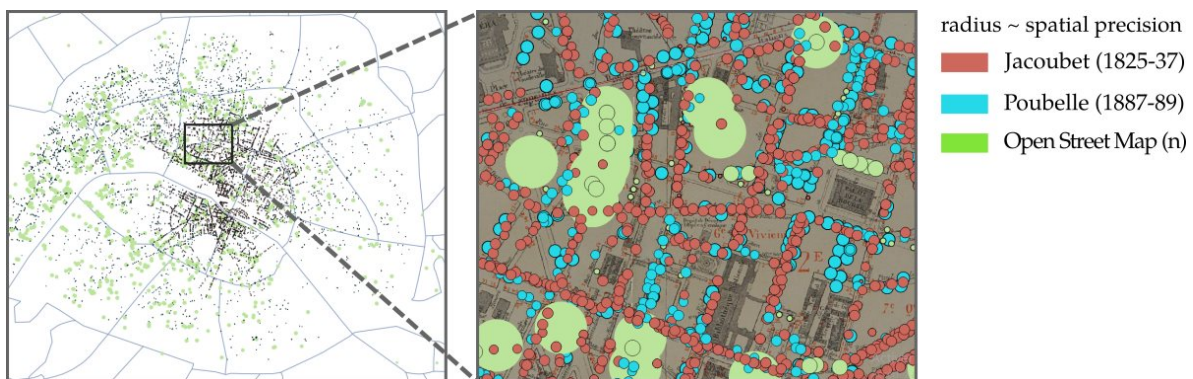


Figure 13. All geocoded historical datasets. Size is proportional to spatial precision.

3.2.1. Manually collected dataset

South Americans

South America immigrants living in Paris in 1926, manually input from census, collected by Elena Monges(EHESS).

Table 1. All geocoded historical datasets facts.

Dataset name	input addresses	found addresses (rough)	second per k addresses
South Americans	13991	13743 (250)	138
Textile	5777	5688 (16)	135
Textile 2	3070	3053 (2)	110
Artists accommodations	13907	10215 (2955)	244
Health administrators	1887	1698 (171)	316
Belle epoque (0.3)	6467	3880(337)	280
Belle epoque (0.5)	6467	6000	351

Textile

Professionals of textile industry in Paris, manually input from the "Almanachs dy Commerce de Paris", from 1793 to 1845, collected by Carole Aubé (EHESS).

Artists accommodations

Addresses of artist studios and artists accommodations between 1791 and 1831, collected by Isabelle Hostein (EHESS) to study their impact on Paris development.

Health administrators

Addresses of health and hygiene administrators in Paris between 1807 and 1919 ([27]), collected by Pascal Cristofoli (EHESS).

3.2.2. Belle epoque

We geocode another set of addresses that are automatically extracted from directory of Paris financial societies between 1871 and 1910. Directories are books referencing address of company (and name and other information). The process of automatic extraction is complex in itself (Project Belle Epoque, [28]), and is out of scope of this article. We only describe it briefly here.

First each page of Paris directories for specific years are photographed. Pictures are then straightened, and information is extracted via an OCR software which has been configured for the directory specific layout. Further rule based processing parse the text into address fields.

As a result of this automatic process, the quality of addresses is often significantly lower than manually edited addresses. Therefore, we test two settings by allowing a greater maximum semantic distance from 0.3 to 0.5 (over 1).

3.3. Collaborative editing

We propose several User Interface for easy geocoding, and collaborative editing of the geocoding results. We informally tested the interfaces and found that they facilitate geocoding, especially for the batch mode.

We also test the collaborative editing in two scenarios. In the first scenario a specialised user geocodes a single address and display the top 3 results corresponding to this address. In the second scenario, 30 random addresses are to be checked/corrected by a regular user

3.3.1. Scenario 1: top 3 results for one address

Using the web application, we geocode the address "10 rue de vaugirard, paris" for the date 1840, and ask for the top 3 results, as shown in first part of illustration 14. A matching building number geohistorical object exists in the three gazetteers extracted from the three maps. Based on the results, we can safely assume that this building number has not changed for the last 2 centuries.



Figure 14. Two usage scenario: First scenario, an expert geocodes an address and analyse the top 3 results through the time. Second scenario: a user check/correct 30 random addresses in Paris.

3.3.2. Scenario 2: check/correct 30 random addresses

In this scenario, a regular user is to check/correct 30 random addresses from the Jacoubet map using the web application. The task is performed quickly, the check and edit of each address is a matter of a few seconds. The main time consuming task is the loading of the background historical map, due to unfortunate hardware limitations. The edit speed is on par with a desktop based edit solution (using QGIS).

4. Discussion

4.1. Geohistorical model

The geohistorical model we propose is designed to be minimal, will keeping all the important sources information. Overall we found it simple enough to be easy to use with all our different historical source (a dozen), yet powerful enough so we never were lacking something essential. The model could however be improved in at least two ways.

4.1.1. Quantifying confidence and degree of knowledge of a source

The first potential improvement would be to integrate a measure of confidence and knowledge over the historical source. This measure would be indicated by an historian with specific knowledge of the source. In fact each historical source could have several confidence depending on the thematic (which is always possible for spatial precision in our current model). For instance, in the Jacoubet Paris map the streets have a high confidence because they were a focus of the map. The building however were most likely updated version of previous maps, thus less trust-worthy. This confidence would be ranked between 0 and 1.

AS an addition, amount of knowledge should also be ranked between 0 and 1, in the same spirit as in the Dempster-Shafer theory. This amount of knowledge would model the possibility for a source to be partial, for instance for a partial map. We feel it is important for an historian to be able to qualify how much the historical source is complete, because it can have major impact on how to deal with the information.

For instance a building missing in a map can be an error of the map (confidence low), or the fact that the map does not cover this building (knowledge low), or the fact that the building was really not existing at the date.

Such an addition to the model (two numbers in $[0..1]$) would be light, an easy to understand and use for historians.

4.1.2. Generalizing temporal definition

Another possible model improvement would be to generalize the temporal aspect definition. We presented fuzzy dates that are represented by a trapezoid probability of existence. This model is

simple to use and understand, but in some case it may prove limited. We could instead let the user define any existence probability function thanks to geometric modeling. For instance an historical source may be supported by three contradictory documents, one stating that the source exists in 1800, the other that the source exists in 1900, and a third that the source does not exist in 1850. In this case a trapezoid function is not generic enough.

4.2. *Geohistorical objects sources*

A large part of the geocoding work depends on the quality and coverage of geohistorical objects that are used for localisation (building numbers, street, city, etc.). In our experiments we use historical source of different natures and of different periods. The historical sources we use allow use to find a very high percentage of addresses for Paris between 1800 and today (2016). Yet the coverage of our historical source could be greatly improved, in term of spatial coverage (adding other cities to geocoder), temporal coverage (reducing gaps in the 20th century), and source quality (better spatial precision).

4.2.1. Better geohistorical sources

The work to exploit historical topographical maps is explained and discussed in detail in [16]. The coverage of street axis is good regarding the available maps, which gives a coherent and solid dataset.

What could be improved is the position of axis (they are not necessarily centered regarding the streets) using an automatic method (optimisation to center the street axis in regard to the detected limits of buildings).

Better axis would indirectly lead to better building number positioning for Alphan map.

4.2.2. More geohistorical sources

We exploit only two historical maps, and only partially, because we do not exploit buildings cadastre.

More objects type

In this article we use several type of geohistorical objects for geocoding : building numbers, streets axis, neighbourhood. We tested other datasets, such as the city limits extracted by the project Geo Historical Data in a collaborative way from the Cassini maps. In fact, a compiled version of city limits (GeoPeople project , [29]) from 1793 to current day created by EHESS has also been tested.

But we could also integrate building cadastre so as to have a building layout associated to an address rather than a point, which would solve an old problem of address points. Indeed, there is currently no consensus as to where a building number address point should be positioned: on the entry door, on the letter box, etc.

More excitingly, in some case more precise data is available, giving the layout of apartment in buildings.

Other topographical maps

We exploit Jacoubet and Alphan maps, yet there are several more to be exploited toward the end of the 19th century, and in the beginning of the 20th century. From the beginning of the 20th century, Paris city administration produced a map per year.

Of course, the main improvement direction would be to add maps of other cities/countries! For France at least, major cities have often been mapped starting from 1900.

cross-referencing historical topographical maps

One way to improve quality of available historical data is by advanced cross referencing. For instance [30,31] proposed a spatio-temporal graph to edit and link historical road axis network between them, thus enabling to transfer information from one map to another.

4.2.3. Non topographical maps

Before the beginning of 19th century, the address system was very different in Paris. In mid 18th century, the address system was in fact that each building would have a specific name (no number, no notion of street name) in its neighbourhood. Our geocoding system has also been designed with this type of addressing but it has not been tested yet.

More generally, this type of indirect localisation is very close to the field of web of knowledge

4.3. Localisation of Historical datasets

Overall the experiments we perform allow use to geocode real historical datasets for several historians, with a very high return rate and quite fast. However from an academic point of view, validating the results of our geocoder is not easy.

4.3.1. Analysing the historical map most used by geocoder

It is interesting to look at what historical sources were the most used for geocoding, although the historical source are chosen based on a complex ranking function.

If we take the example of the over 10k geocoded addresses from the "Artists accommodations" dataset, we could expect all of the results to be drawn from the Jacoubet map, as the dataset is between 1793 and 1836, and the Jacoubet map is also in this range. Yet, analysing the results shows that if Jacoubet was used for 80% of the addresses, Alphand was used for 15%, although the map comes 30 years after. More surprisingly, the OpenStreetMap current data is still used for 5% of addresses, although it is about 2 centuries after the dataset.

Similar analysis on other datasets show similarly that all maps are always used, with of course a focus on the temporally closest map.

We think that this results are explained by the fact that historical maps miss some information, contains error, and do not have the same geographical coverage.

4.3.2. Geocoding qualification and quality measures

Modern geocoders are evaluated by how often they find a localisation, and how precise is the localisation they return (see [32] for instance).

For historical geocoding, both measures are difficult. Indeed, the percentage of found addresses directly depends on the available historical sources completeness. Moreover, there is no way to guarantee that the input geocoding query (address and date) actually existed.

Estimating the spatial quality of the result is similarly difficult, because there are no ground truth to compare the results to!

4.3.3. Scalability

The main design choice of our geocoding architecture is to use a flat model for the address (an address is any set of characters), as opposed to current geocoder which are highly hierarchical (an address refers to a street, that refers to a neighbourhood, etc.). This modelling choice gives the freedom that is necessary for data as incomplete as the historical ones, but also comes with a tradeoff regarding scaling capabilities.

Indeed, for strongly hierarchical data, it is possible to have separate databases for each city for instance, thus preventing one database to grow too much, and ensuring a nice scaling capability.

This is not however the case with our architecture. By using database indexes, we can theoretically guarantee a fast geocoding time for up to few dozen of millions of geohistorical object used as sources. The main bottleneck in this case is not the temporal aspect (it relies on PostGIS geometry, which enable multiple theoretical solution for scaling), but the semantic aspects (i.e. the address string itself). To scale over dozens of million of addresses, specific architectures may be used to deal with the semantic search, for instance distributed database (database sharding), in a similar spirit to the current software Elastic Search. We stress however that given the current available amount of historical sources, such scaling problem should not be an issue before a long time.

4.4. Collaborative editing

We propose several ways to use the geocoding capabilities in an easy way through web based User Interfaces. As we proposed prototypes, the experiments are proofs of concepts. For a real validation, a complete user study would be required, which is outside of the scope of this article.

4.4.1. improving UI

The User interface could however be improved. First a time slider would be helpful in the scenario where several candidates are displayed. Second the underlying displayed historical map is paramount for the user to be able to check and edit the results. For the moment the user has to set the map he wants, yet the more appropriate map could be chosen automatically by temporal proximity for instance. Last we feel that in the current interface the map can become easily cluttered when too many results are displayed at the same time, because the labels occupy a large amount of screen space. We could better use the current clustering system that group the addresses together when they are close on the screen.

4.4.2. Integrating user correction into historical sources

In collaborative editing, edit come from untrusted sources. Validating edits and solving conflicts is then a classical problem. In our prototypes every user edit is potentially used by the geocoder (they are added to a dedicated gazetteer). We could use a voting scheme where edits are only taken into account when a sufficient number of user have made them. However, we stress that due to the number of data to edit (several hundred thousands building numbers), we prefer to rely on the user benevolence, by considering that user spending time editing centuries old historical data are committed to accurate editing.

4.5. Beyond geocoding

5. Conclusion

This article tackles the historical geocoding problem. The historical aspects brings major complications. The main difficulties comes from the nature of historical data (uncertainty, fuzzy date, precision, sparseness), which prevents the use of current-address geocoding methods based on strong hierarchical modelling. Instead, we propose a historical geocoding system based on a sound model of geohistorical objects. This model is designed to cover the minimal features, and by its generality, modularity, and open source nature, can easily be extended to feat other historical sources. We integrate into the database geohistorical objects from historical sources that have been coherently georeferenced and edited to form gazetteers. Geocoding and address at a given time is then a matter of finding the best matching geohistorical object in the gazetteers, if any. Our simple, coherent historical geocoding system tested on several real-life datasets collected by historians can be easily used for other place/time/type of localisations. We integrate into the geocoder diverse historical sources covering two century for the city of Paris. The geocoder is able to localise a large percentage of addresses with a fast speed of about 200ms per address. We propose a prototype

of web-based User Interface that demonstrate the interest of collaborative editing of localisation of addresses, and help use of geocoding for historians.

Supplementary Materials: All the code, and additionnal documentation is available on the project website : <https://github.com/Geohistoricaldata>

Acknowledgments: Thanks to the Belle Epoque project and Angelo Riva and Thierry Géraud, the Institut Louis Bachelot for funding. Thanks to historians who contributed to creating the datasets, especially Benoit Costes for the edit of Alphand map.

Bibliography

1. Davis, C.A.; Fonseca, F.T. Assessing the Certainty of Locations Produced by an Address Geocoding System. *Geoinformatica* **2007**, *11*, 103–129.
2. Goldberg, D.W.; Wilson, J.P.; Knoblock. From Text to Geographic Coordinates: The Current State of Geocoding. *Journal of the Urban and Regional Information Systems Association* **2007**, *19*, 33–46.
3. Hutchinson, M.J.; Veenendaal, B. An agent-based framework for intelligent geocoding. *Applied Geomatics* **2013**, *5*, 33–44.
4. Roongpiboonsopit, D.; Karimi, H.A. Comparative Evaluation and Analysis of Online Geocoding Services. *International Journal of Geographical Information Science* **2010**, *24*, 1081–1100.
5. Clough, P.; Tang, J.; Hall, M.M.; Warner, A. Linking archival data to location: A case study at the UK national archives. *Aslib Proceedings* **2011**, *63*, 127–147.
6. Mostern, R.; Johnson, I. From named place to naming event: creating gazetteers for history. *International Journal of Geographical Information Science* **2008**, *22*, 1091–1108, [<http://dx.doi.org/10.1080/13658810701851438>].
7. Smith, D.A.; Crane, G. Disambiguating Geographic Names in a Historical Digital Library. *International Conference on Theory and Practice of Digital Libraries*. Springer, 2001, pp. 127–136.
8. St-Hilaire, M.; Moldofsky, B.; Richard, L.; Beaudry, M. Geocoding and Mapping Historical Census Data: The Geographical Component of the Canadian Century Research Infrastructure. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* **2007**, *40*, 76–91.
9. Goodchild, M.F. Citizens as sensors: the world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221.
10. Heipke, C. Crowdsourcing geospatial data. *{ISPRS} Journal of Photogrammetry and Remote Sensing* **2010**, *65*, 550 – 557. *{ISPRS} Centenary Celebration Issue*.
11. Haklay, M., Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. In *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*; Sui, D.; Elwood, S.; Goodchild, M., Eds.; Springer Netherlands: Dordrecht, 2013; pp. 105–122.
12. Fomel, S.; Claerbout, J.F. Guest Editors' Introduction: Reproducible Research. *Computing in Science Engineering* **2009**, *11*, 5–7.
13. Aruliah, D.A.; Brown, C.T.; Hong, N.P.C.; Davis, M.; Guy, R.T.; Haddock, S.H.D.; Huff, K.; Mitchell, I.; Plumbley, M.D.; Waugh, B.; White, E.P.; Wilson, G.; Wilson, P. Best Practices for Scientific Computing. *CoRR* **2012**, *abs/1210.0530*.
14. Wilson, G.; Bryan, J.; Cranston, K.; Kitzes, J.; Nederbragt, L.; Teal, T.K. Good Enough Practices in Scientific Computing. *ArXiv e-prints* **2016**, [[arXiv:cs.SE/1609.00037](https://arxiv.org/abs/1609.00037)].
15. Marwick, B. Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation. *Journal of Archaeological Method and Theory* **2016**, pp. 1–27.
16. Duménieu, B. Un Système d'information Géographique Pour Le Suivi d'objets Historiques Urbains à Travers l'espace et Le Temps. PhD thesis, Ecole des Hautes Etudes en Sciences Sociales, 2015.
17. Armstrong, M.P. Temporality in spatial databases. *GIS/LIS'88*, 1988, pp. 880–889.
18. Herrault, P.A.; Sheeren, D.; Fauvel, M.; Monteil, C.; Paegelow, M. A comparative study of geometric transformation models for the historical "map of france" registration. *Geographia Technica* **2013**, pp. pp–34.
19. Fabbri, R.; Kimia, B. 3D curve sketch: Flexible curve-based stereo reconstruction and calibration. *Computer Vision and Pattern Recognition (CVPR)*, 2010 IEEE Conference on. IEEE, 2010, pp. 1538–1545.
20. Cléri, I.; Pierrot-Deseilligny, M.; Vallet, B. Automatic Georeferencing of a Heritage of old analog aerial Photographs. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2014**, *2*, 33.

21. Bitelli, G.; Cremonini, S.; Gatta, G. Ancient map comparisons and georeferencing techniques: a case study from the Po River Delta (Italy). *E-perimetron* **2009**, *4*, 221–228.
22. Boutoura, C.; Livieratos, E. Some fundamentals for the study of the geometry of early maps by comparative methods. *e-Perimetron* **2006**, *1*, 60–70.
23. De Runz, C.; Desjardin, E.; Piantoni, F.; Herbin, M. Anteriority index for managing fuzzy dates in archæological GIS. *Soft Computing* **2010**, *14*, 339.
24. Kauppinen, T.; Mantegari, G.; Paakkari, P.; Kuittinen, H.; Hyvönen, E.; Bandini, S. Determining relevance of imprecise temporal intervals for cultural heritage information retrieval. *International journal of human-computer studies* **2010**, *68*, 549–560.
25. Perret, J.; Gribaudo, M.; Barthelemy, M. Roads and cities of 18th century France. *Scientific Data* **2015**, *2*, 10.1038/sdata.2015.48.
26. Noizet, H.; Bove, B.; Costa, L. *Paris de Parcelles En Pixels*; Presses universitaires de Vincennes, 2013.
27. Gribaudo, M.; Magaud, J. L'action publique et ses administrateurs dans les domaines sanitaires et social en France, 1800 à 1900, 1999.
28. Lazzara, G.; Levillain, R.; Géraud, T.; Jacquélet, Y.; Marquegnies, J.; Crépin-Leblond, A. The SCRIBO Module of the Olena Platform: A Free Software Framework for Document Image Analysis. Proceedings of the 11th International Conference on Document Analysis and Recognition (ICDAR); , 2011; pp. 252–258.
29. Plumejeaud-Perreau, C.; Grosso, E.; Parent, B. Dissemination and Geovisualization of Territorial Entities History. *Journal of Spatial Information Science* **2014**, *8*, pp. 73–93.
30. Costes, B.; Perret, J.; Bucher, B.; Gribaudo, M. An aggregated graph to qualify historical spatial networks using temporal patterns detection. 18th AGILE International Conference on Geographic Information Science, 2015.
31. Costes, B. Vers la construction d'un référentiel géographique ancien. Un modèle de graphe agrégé pour intégrer, qualifier et analyser des réseaux géohistoriques. PhD thesis, Université Paris-Est, 2016.
32. Zimmerman, D.L.; Fang, X.; Mazumdar, S.; Rushton, G. Modeling the Probability Distribution of Positional Errors Incurred by Residential Address Geocoding. *International Journal of Health Geographics* **2007**, *6*, 1.